

Accuracy of Wearable Heart Rate Monitors under Static and High-Intensity Activity Conditions

János Dávid Balogh^a, Attila Adamkó^b

^aFaculty of Informatics, University of Debrecen
davidkm2@mailbox.unideb.hu

^bFaculty of Informatics, University of Debrecen
adamko.attila@inf.unideb.hu

Abstract

The spread of wearable devices has created new opportunities for continuous monitoring of users' health. However, it remains an open question to what extent the physiological data provided by commercial smartwatches can be considered accurate and clinically interpretable. The aim of this study is to compare heart rate, heart rate variability (HRV), photoplethysmography (PPG) measurements, as well as step count and sleep parameters with reference measuring devices in different activity environments, and to assess the applicability of currently available devices for health monitoring and sports health.

During the study, we compared commercial smartwatches from four different manufacturers with standard reference devices in controlled laboratory and everyday environments. Based on our results, heart rate in a static environment showed a high correlation with the reference ($r > 0.92$), while during intense physical activity the error rate increased significantly (MAE = 10–18 bpm). For HRV data, variability and consistency between devices were lower, while sleep duration estimation showed moderate accuracy and sleep stage determination showed poor reliability.

The results suggest that smartwatches are suitable for detecting longer-term health trends, but their use for clinical decision support purposes is currently limited. The study also highlights the need to improve validation practices, strengthen algorithmic transparency, and enhance interoperability.

Problem 1 (HR accuracy (stable environment)). *Commercially available wearable devices are increasingly used to monitor physiological parameters. However, the accuracy of heart rate measurement remains controversial, especially when compared to reference devices. In static environments, the temporal dynamics of physiological variables are low, which favors the accuracy of photoplethysmographic sensors. Accordingly, in static activity environments, HR data from wearable devices are expected to show small deviations compared to clinically validated reference devices.*

Definition 2. A static environment is a measurement situation in which the subject’s movement intensity is minimal and the change in heart rate shows low temporal dynamics.

Definition 3. Any instrument that provides HR measurement at a clinical level of accuracy and whose validation is documented in published literature can be considered a reference instrument [3].

Definition 4. The measurement error can be defined as the difference between the HR measured by the wearable device and the HR value of the reference device:

$$e(t) = HR_{\text{wearable}}(t) - HR_{\text{ref}}(t)$$

Definition 5. In the case of accurate measurement, the expected value of the measurement error is close to zero, and its standard deviation remains within a predefined clinical tolerance [4].

Theorem 6. *In a static environment, the average HR values measured by the smartwatch do not show significant differences from the HR value of the reference device: $H_0 : \mu_{\text{wearable}} = \mu_{\text{ref}}$ Statistical testing of the hypothesis can be performed by paired sample comparisons (paired t-test or Wilcoxon sign-rank), while clinical agreement can be characterized by Bland–Altman analysis [1].*

Corollary 7. *HR measurements from wearable devices are suitable for tracking baseline parameters and trend detection in static environments. However, measurement accuracy in dynamic environments cannot be generalized and requires separate validation. This conclusion is consistent with previous studies that found the accuracy of wearable devices to be adequate in static situations, while showing significant differences during high-intensity physical activity [2].*

Problem 8 (Accuracy of HR measurement during intense physical activity). *Heart rate measurements from wearable devices during intense physical activity may deviate significantly from clinically validated reference devices. PPG-based sensors are sensitive to skin tension, motion artifacts, and blood flow dynamics, which can lead to increased measurement error. Based on the problem definition, the accuracy of HR measurements in intense activity environments may be limited, which affects the clinical or decision-support applicability of wearable devices.*

Definition 9. Intense physical activity refers to forms of exercise during which the heart rate increases significantly and dynamic changes in the body are rapid, such as running, interval training, or fast cycling.

Definition 10. A clinically validated HR monitor that measures heart rate with high accuracy (99%), such as a chest strap-based ECG sensor, can be considered a reference device [3].

Definition 11. The measurement error is the difference between the HR measured by the wearable device and the reference device:

$$e(t) = HR_{\text{wearable}}(t) - HR_{\text{ref}}(t)$$

Definition 12. A significant difference can be considered if the average heart rate measured by the wearable device is statistically different from the value measured by the reference device. Formally:

$$H_0 : \mu_{\text{wearable}} = \mu_{\text{ref}} \quad \text{vs.} \quad H_1 : \mu_{\text{wearable}} \neq \mu_{\text{ref}},$$

where H_0 is the null hypothesis, H_1 is the alternative hypothesis, and μ_{wearable} and μ_{ref} denote the corresponding mean heart rate values. To check significance, a paired-sample t-test or a nonparametric Wilcoxon sign-rank test can be used, taking into account the statistical characteristics of measurement errors.

Theorem 13. *During intense physical activity, HR values measured by wearable devices differ significantly from HR values measured by the reference device:*

$$H_1 : \mu_{\text{wearable}} \neq \mu_{\text{ref}}$$

Significance is checked using paired t-test or Wilcoxon sign-rank test. Bland–Altman analysis and RMSE calculation can be used to characterize the extent of differences [1]

Corollary 14. *HR measurements from wearable devices during intense physical activity do not provide the same accuracy as in a static environment. As a result, trend detection and sports performance monitoring are still possible, but further validation and calibration are required for clinical decision support or critical healthcare applications.*

References

- [1] J. M. BLAND, D. G. ALTMAN: *Statistical methods for assessing agreement between two methods of clinical measurement*, The Lancet 327.8476 (1986), pp. 307–310, DOI: [10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8).
- [2] S. GILLINOV, M. ETIWY, et AL.: *Variable accuracy of wearable heart rate monitors during aerobic exercise*, Medicine & Science in Sports & Exercise 49.8 (2017), pp. 1697–1703, DOI: [10.1249/MSS.0000000000001284](https://doi.org/10.1249/MSS.0000000000001284).
- [3] J. PARAK, I. KORHONEN: *Evaluation of wearable consumer heart rate monitors based on photoplethysmography*, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Chicago, IL, USA, 2014, pp. 3670–3673, DOI: [10.1109/EMBC.2014.6944419](https://doi.org/10.1109/EMBC.2014.6944419).
- [4] M. P. WALLEN, S. R. GOMERSALL, S. E. KEATING, U. WISLØFF, J. S. COOMBES: *Accuracy of heart rate watches: Implications for weight management*, PLoS ONE 11.5 (2016), e0154420, DOI: [10.1371/journal.pone.0154420](https://doi.org/10.1371/journal.pone.0154420).